(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号

特開平10－214258

(43)公開日　平成10年(1998)８月11日

(51)Int.Cl.[6]		識別記号	ＦＩ		
Ｇ０６Ｆ	15/16	３８０	Ｇ０６Ｆ	15/16	３８０Ｚ
	3/16	３２０		3/16	３２０Ｈ
		３３０			３３０Ｋ
	13/00	３５１		13/00	３５１Ｇ
		３５７			３５７Ｚ

審査請求　未請求　請求項の数４　ＦＤ　(全 10 頁)　最終頁に続く

(21)出願番号　特願平9－28409

(22)出願日　平成９年(1997)１月28日

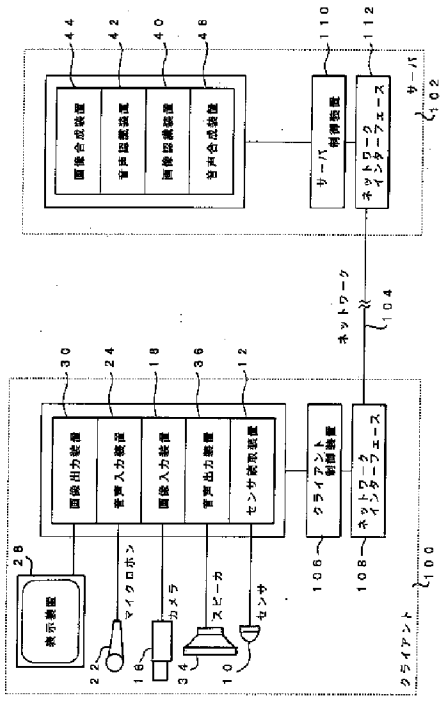
(71)出願人　000004329
日本ビクター株式会社
神奈川県横浜市神奈川区守屋町３丁目12番地

(72)発明者　丸山　俊弘
神奈川県横浜市神奈川区守屋町３丁目12番地　日本ビクター株式会社内

(74)代理人　弁理士　梶原　康稔

(54)【発明の名称】　データ処理システム

(57)【要約】
【課題】　ハードウエア資源を効率的に活用でき、空間的な制約を低減することができるデータ処理システムを提供する。
【解決手段】　高速の処理や大規模な記憶容量を必要とする画像認識・合成や音声認識・合成などの装置40～46は、サーバ102側にのみ用意される。クライアント100側では、画像や音声の入出力が行われる。クライアント100側で入力された画像データや音声データは、ネットワーク104を通じてサーバ102側に転送され、サーバ側で認識処理される。そして、認識処理後の画像データ及び音声データ，あるいはサーバ側で合成処理された画像データ及び音声データは、ネットワーク104を通じてクライアント側に転送され、クライアント側で表示あるいは再生される。



1

【特許請求の範囲】

【請求項１】　クライアントは、

画像の入力を行って画像データを得る画像入力手段，音声の入力を行って音声データを得る音声入力手段，画像データに基づいて画像を表示する画像出力手段，音声データに基づいて音声を再生する音声出力手段，ネットワークとの接続を行うネットワークインターフェース，前記画像入力手段及び前記音声入力手段によってそれぞれ得られた画像データ及び音声データをサーバ側に転送するとともに、サーバ側から転送された画像データ及び音声データを前記画像出力手段及び前記音声出力手段にそれぞれ供給する動作を制御するクライアント制御手段，を備えており、

サーバは、

クライアント側から転送された画像データを認識する画像認識手段，クライアント側から転送された音声データを認識する音声認識手段，画像データを合成する画像合成手段，音声データを合成する音声合成手段，ネットワークとの接続を行うネットワークインターフェース，クライアント側から転送された画像データ及び音声データを前記画像認識手段及び前記音声認識手段にそれぞれ供給するとともに、前記画像合成手段及び前記音声合成手段によって合成された画像データ及び音声データをクライアント側にそれぞれ転送する動作を制御するサーバ制御手段，を備えており、

これらクライアント及びサーバをネットワークによって接続したことを特徴とするデータ処理システム。

【請求項２】　前記クライアントを前記ネットワークに複数接続するとともに、前記サーバからのデータ転送先のクライアントを識別するクライアント識別手段を備えたことを特徴とする請求項１記載のデータ処理システム。

【請求項３】　前記サーバを前記ネットワークに複数接続するとともに、前記サーバの処理状況に応じてクライアントからのデータ転送先のサーバを選択するサーバ選択手段を備えたことを特徴とする請求項１又は２のいずれかに記載のデータ処理システム。

【請求項４】　画像及び音声の処理対象を検知するためのセンサ手段を備えたことを特徴とする請求項１，２又は３のいずれかに記載のデータ処理システム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】この発明は、画像データ及び音声データに対して所望の処理を行うデータ処理システムに関するものである。

【０００２】

【背景技術】画像データや音声データに対して認識や合成などの各種の処理を行うデータ処理システムとしては、例えば、図３，図４に示すものがある。この例は、事前に登録された特定の人物がシステムの前に現れ、その音声と顔面を入力して人物を特定し、登録済みのメッセージと画像を再生出力するようなシステムである。図３にはシステムの構成例が示されており、これを更にブロック化したものが図４に複数示されている。これらの図において、人物の到来などのセンサ読取は、赤外線センサ，振動センサ，音声センサなどのセンサ１０及びセンサ読取装置１２によって行われ、これらによってセンサ読取部１４が構成されている。

2

【０００３】人物の画像入力とデジタル化データへの変換は、カメラ１６及び画像入力装置１８によって行われ、これらによって画像入力部２０が構成されている。人物の音声入力とデジタル化データへの変換はマイクロホン２２及び音声入力装置２４によって行われ、これらによって音声入力部２６が構成されている。デジタル化された画像データに基づく画像出力は、表示装置２８及び画像出力装置３０によって行われ、これらによって画像出力部３２が構成されている。デジタル化された音声データに基づく音声出力は、スピーカ３４及び音声出力装置３６によって行われ、これらによって音声出力部３８が構成されている。

【０００４】これらの各装置によって得られた人物の顔面画像や音声のデータは、画像認識装置４０あるいは音声認識装置４２に供給され、ここで認識のための解析処理が行われる。また、合成すべき画像データや音声データがある場合には、それらに基づいて画像合成装置４４や音声合成装置４６で画像や音声が合成される。画像認識の例としては、人物判定や動作解析などがある。音声認識の例としては、特定話者の判定や会話の解析などがある。画像合成の例としては、３次元画像のレンダリングや動画データの生成などがある。音声合成の例としては、任意の音色による音声合成などがある。

【０００５】合成後の画像は画像出力装置３０に供給され、更に表示装置２８に表示される。また、合成後の音声は、音声出力装置３６に供給され、更にマイクロホン２２で再生される。これらの処理は、制御装置４８によって制御される。

【０００６】以上のような画像や音声の入出力装置や認識合成装置は、図４に示すように各システム毎にそれぞれ設けられる。すなわち、各システム毎に独立して、センサ読取，画像や音声の入力，認識合成処理，出力がそれぞれ行われる。

【０００７】

【発明が解決しようとする課題】ところで、一般的に、画像の認識や音声の認識では、膨大なデータが高速で処理され、あるいは記憶される。このため、高速のＣＰＵやデータ処理用のＤＳＰ，専用ハードウエア，大容量の記憶装置などが必要となる。同様に表示する画像を生成する画像合成や、音声を生成する音声合成についても、高速のＣＰＵや高速データ処理用のＤＳＰ，その他の専用ハードウエアを必要とする。画像や音声の認識や合成

は、それらに依存して実現されている。

【０００８】ところが、それらの高速のＣＰＵ，データ処理用のＤＳＰ，専用のハードウエア，大容量の記憶装置は、いずれも高価であり、システムのコストアップを招くとともに、装置構成も複雑となる。従って、上述した背景技術のように、各システム毎に独立して画像や音声の認識装置あるいは合成装置を備える構成とすると、各システムをワークステーションなどの高性能なコンピュータで構成する必要がある。従って、各システムのコストが非常に高価になってしまう。特に、図４に示したように、同じ機能を備えたシステムを複数用意するような場合、端末のコストが高いために台数に比例してコストがかかり、多数のシステムを用意することは膨大なコストを必要とすることになる。また、システムの小型化，省スペース化という点からも好ましいとは言えない。

【０００９】この発明は、以上の点に着目したもので、その目的は、ハードウエア資源を効率的に活用できる低コストのデータ処理システムを提供することである。他の目的は、空間的な制約を低減することができるデータ処理システムを提供することである。

【００１０】

【課題を解決するための手段】前記目的を達成するため、この発明では、クライアント（100）側は、画像の入力を行って画像データを得る画像入力手段（16,18），音声の入力を行って音声データを得る音声入力手段（22,24），画像データに基づいて画像を表示する画像出力手段（28,30），音声データに基づいて音声を再生する音声出力手段（34,36），ネットワーク（104）との接続を行うネットワークインターフェース（108），前記画像入力手段及び前記音声入力手段によってそれぞれ得られた画像データ及び音声データをサーバ側に転送するとともに、サーバ側から転送された画像データ及び音声データを前記画像出力手段及び前記音声出力手段にそれぞれ供給する動作を制御するクライアント制御手段（106），を備えている。

【００１１】サーバ（102）側は、クライアント側から転送された画像データを認識する画像認識手段（40），クライアント側から転送された音声データを認識する音声認識手段（42），画像データを合成する画像合成手段（44），音声データを合成する音声合成手段（46），ネットワークとの接続を行うネットワークインターフェース（112），クライアント側から転送された画像データ及び音声データを前記画像認識手段及び前記音声認識手段にそれぞれ供給するとともに、前記画像合成手段及び前記音声合成手段によって合成された画像データ及び音声データをクライアント側にそれぞれ転送する動作を制御するサーバ制御手段（110），を備えている。

【００１２】そして、これらクライアント及びサーバがネットワークによって接続されたことを特徴とする。

【００１３】主要な形態によれば、前記クライアントが前記ネットワークに複数接続されるとともに、前記サーバからのデータ転送先のクライアントを識別するクライアント識別手段（200,202）が備えられる。あるいは、前記サーバが前記ネットワークに複数接続されるとともに、前記サーバの処理状況に応じてクライアントからのデータ転送先のサーバを選択するサーバ選択手段（300,302）が備えられる。また、画像及び音声の処理対象を検知するためのセンサ手段（10,12）が備えられる。

【００１４】他の主要な形態では、音声の入力をデジタル化された入力音声データに変換する音声入力装置，画像の入力をデジタル化された入力画像データに変換する画像入力装置，デジタル化された出力音声データを再生する音声出力装置，デジタル化された出力画像データを表示する画像出力装置，人物の存在を検知するセンサ，ネットワークとの接続を行うネットワークインターフェース，クライアントの動作を制御するクライアント制御装置をクライアントが備えている。また、ネットワークを介して転送されてきた入力音声データを解析・認識を行う音声認識装置，ネットワークを介して転送されてきた入力画像データを解析・認識を行う画像認識装置，出力音声データの合成を行う音声合成装置，出力画像データの合成を行う画像合成装置，ネットワークを介して前記クライアントと入力音声データ、入力画像データ、出力音声データ、出力画像データの転送を行うサーバ側ネットワークインターフェース，サーバ側の動作を制御するサーバ制御装置をサーバが備えている。そして、これらのクライアントとサーバがネットワークによって結ばれている。

【００１５】本発明によれば、クライアント側では、画像や音声の入出力が行われる。クライアント側で入力された画像データや音声データは、ネットワークを通じてサーバ側に転送され、サーバ側で認識処理される。そして、認識処理後の画像データ及び音声データ，あるいはサーバ側で合成処理された画像データ及び音声データは、ネットワークを通じてクライアント側に転送され、クライアント側で表示あるいは再生される。高速の処理や大規模な記憶容量を必要とする画像認識・合成や音声認識・合成などの装置をサーバ側にのみ用意すればよいので、ハードウエア資源を効率的に活用することができる。この発明の前記及び他の目的，特徴，利点は、以下の詳細な説明及び添付図面から明瞭になろう。

【００１６】

【発明の実施の形態】以下、発明の実施の形態について、実施例を参照しながら詳細に説明する。

【００１７】

【実施例１】最初に、図１及び図２（Ａ）を参照して実施例１を説明する。なお、上述した背景技術に対応する要素には同一の符号を用いる。このシステムは、図２（Ａ）に全体を示すように、クライアント（端末）１０

０とサーバ１０２がネットワーク１０４によって接続された構成となっている。図１には、各部の詳細な構成が示されている。同図において、上述したセンサ読取装置１２，画像入力装置１８，音声入力装置２４，画像出力装置３０，音声出力装置３６は、クライアント制御装置１０６によって動作が制御されるようになっている。クライアント１００には以上の各装置が設けられている。そして、ネットワーク１０４に対するクライアント１００の接続は、ネットワークインターフェース１０８によって行われるように構成されている。

【００１８】一方、上述した画像認識装置４０，音声認識装置４２，画像合成装置４４，音声合成装置４６は、いずれもサーバ１０２に設けられている。これらの各装置は、サーバ制御装置１１０によって動作が制御されている。そして、ネットワーク１０４に対するサーバ１０２の接続は、ネットワークインターフェース１１２によって行われるように構成されている。

【００１９】以上のように、画像や音声の入出力部がクライアント１００側に設けられており、画像や音声の認識合成装置はサーバ１０２側に設けられている。そして、クライアント１００とサーバ１０２とがネットワーク１０４で接続された構成となっている。

【００２０】次に、全体の動作を説明する。なお、上述した背景技術と同様に、事前に登録された特定の人物がクライアント１００の前に現れてその音声と顔面の画像と音声を入力するとともに、サーバ１０２でその人物を特定し、更に登録済みの画像や音声メッセージをクライアント１００で再生出力する処理を行う場合を説明する。

【００２１】クライアント１００に対する人物の接近は、センサ１０によって検知される。すると、画像入力部２０のカメラ１６，画像入力装置１８によって人物の顔面が撮影され、画像データとして取り込まれる。更に、人物が音声を発したときは、その音声が音声入力部２６のマイクロホン２２，音声入力装置２４によって入力され、音声データとして取り込まれる。これらの取り込まれた画像データと音声データは、クライアント制御装置１０６により、ネットワークインターフェース１０８を通じてネットワーク１０４に供給され、サーバ１０２側に転送される。

【００２２】サーバ１０２では、ネットワーク１０４を経て転送されてきたクライアント１００からのデータが、ネットワークインターフェース１１２によって取り込まれる。取り込まれたデータのうち、画像データはサーバ制御装置１１０によって画像認識装置４０に供給され、音声データはサーバ制御装置１１０によって音声認識装置４２に供給される。

【００２３】画像認識装置４０では、予め用意された画像認識アルゴリズムに基づいて画像データが解析される。例えば、「特徴抽出マッチング」，「パターンマッチング」などの手法によって、その解析が行われる。そして、解析結果と、予めサーバ１０２側の記憶装置（図示せず）に登録されている人物の顔面画像データとが比較され、一致するかもしくは最もデータが近似する人物を選択するなどの手法によって、該当する人物が識別される。

【００２４】他方、音声認識装置４２では、予め用意された音声認識アルゴリズムに基づいて音声データが解析される。例えば、「ＤＰ（Dynamic Programing）マッチング」，「HMM（隠れマルコフモデル）」などの手法によって、その解析が行われる。そして、解析結果と、記憶装置に予め登録されている単語や熟語などの音声データとが比較され、一致するかもしくは最もデータが近似するものを選択するなどの手法によって、該当する言葉が識別される。

【００２５】以上のような認識結果は、サーバ制御装置１１０によりネットワークインターフェース１１２，ネットワーク１０４を介してクライアント１００側に転送される。クライアント制御装置１０６は、ネットワーク１０４を通じて送られてきた画像認識結果と音声認識結果を、ネットワークインターフェース１０８を介して受け取る。続いて、合成すべき画像データや音声データがある場合には、クライアント制御装置１０６により、ネットワークインターフェース１０８，ネットワーク１０４を介してサーバ１０２側に合成すべき画像データと音声データが転送される。なお、合成すべき画像データは、実際の画像データではなく、例えば予め決められた画像を示す番号データなどである。合成すべき音声データも、実際の音声波形をデジタル化したデータではなく、例えばテキストなどの文字列情報又は予め決められた定形のメッセージ番号データなどである。

【００２６】サーバ制御装置１１０は、ネットワーク１０４を通して受信した合成すべき画像データや音声データを、ネットワークインターフェース１１２を介して受け取る。そして、サーバ制御装置１１０により、合成すべき画像データは画像合成装置４４に、合成すべき音声データは音声合成装置４６にそれぞれ供給される。

【００２７】画像合成装置４４では、入力された合成すべき画像データに基づいて、実際の合成画像データが生成される。例えば、動画が必要な場合には、複数の画像から動画データが生成される。画像を合成する際に、「レンダリング」などのアルゴリズムを用いてその場で描画を行なってもよい。また、サーバ１０２側に設けられた記憶装置に蓄積された画像データや動画データを読み出すのみでもよい。音声合成装置４６では、入力された合成すべき音声データに基づいて、実際の合成音声データ（波形データ）が生成される。例えば、「分析音合成」，「規則音合成」などのアルゴリズムによって音声が合成される。また、記憶装置に蓄積された音声データ読み出すのみでもよい。

【００２８】画像合成装置４４，音声合成装置４６によって合成されたデータは、サーバ制御装置１１０により、ネットワークインターフェース１１２，ネットワーク１０４を介してクライアント１００側へ送られる。クライアント１００側では、クライアント制御装置１０６により、転送された合成画像データが画像出力装置３０に供給される。そして、画像出力装置３０によって、合成画像データに基づく画像が表示装置２８に表示される。また、転送された合成音声データは、クライアント制御装置１０６により音声出力装置３６に供給される。そして、音声出力装置３６によって、合成音声データに基づく音声がスピーカ３４から出力される。

【００２９】以上のように、実施例１によれば、クライアント１００側に画像や音声の入出力装置が設けられる。そして、画像や音声の認識装置や合成装置は、サーバ１０２側に設けられる。クライアント１００では、画像や音声の入出力のみが行われ、画像や音声のデータはネットワーク１０４を通じてサーバ１０２側に転送される。そして、高速な処理が要求される画像や音声の認識や合成処理はサーバ１０２上で蓄積データを利用して行なわれる。そして、その処理結果が、ネットワーク１０４を通じてクライアント１００に転送される。

【００３０】このため、システム全体としてみると、クライアントとサーバを別個に配置することが可能となり、設置スペースに対する制約が緩和されるようになる。また、本実施例によれば、クライアント側に人物の存在を検知するためのセンサが設けられているので、画像や音声による誤判定を避けることができるという利点もある。

【００３１】

【実施例２】次に、図２（B）を参照しながら実施例２について説明する。この例は、サーバ１０２に、ネットワーク１０４を通じてクライアント１００Ａ，１００Ｂをそれぞれ接続した構成となっている。クライアント１００Ａ，１００Ｂでは、実施例１と同様に画像や音声の入出力のみが行われる。そして、それら画像や音声の認識合成処理は、実施例１と同様にサーバ１０２で行われる。すなわち、複数のクライアントからネットワークを通じて転送されてくる画像データや音声データに対する処理が、１台のサーバで行われる。

【００３２】クライアント間におけるデータの認証は、転送データにクライアントを識別するための識別データを付加するなどの方法で可能である。例えば、クライアント１００Ａ，１００Ｂに識別データ付加部２００がそれぞれ設けられる。一方、サーバ１０２にはクライアント識別部２０２が設けられる。データの送り元のクライアント１００Ａ又は１００Ｂでは、識別データ付加部２００により識別データを転送データに付加してサーバ１０２側に転送する。サーバ１０２側では、クライアント識別部２０２（又はサーバ制御装置１１０）において識別データを記憶する。転送データの処理結果をクライアント側に返す場合には、記憶した識別データを参照し、クライアント識別部２０２で該当するクライアントを識別して処理結果が転送される。このようにして、複数のクライアントからのアクセスに対応できるようになる。

【００３３】このように、本実施例によれば、各クライアントでは、画像や音声の入出力のみが行われ、画像や音声のデータはネットワークを通じてサーバ側に転送される。そして、高速な処理が要求される画像や音声の認識や合成処理，データの蓄積は、複数のクライアントに共通に設けられたサーバ上で行なわれ、画像や音声の処理結果が該当するクライアントに転送される。このため、高性能のＣＰＵやＤＳＰ，大容量のメモリが各クライアントに共通に設けられるようになり、全体としてコストが削減できる。また、複数のクライアントをネットワークを通じてサーバに接続する構成となっているため、更にクライアントが必要となった場合にもそれをネットワークに接続するのみでよく、ハードウエア資源を効率的に活用することが可能である。

【００３４】

【実施例３】次に、図２（C）を参照しながら実施例３について説明する。この例は、クライアント１００に、ネットワーク１０４を通じてサーバ１０２Ａ，１０２Ｂをそれぞれ接続した構成となっている。クライアント１００では、実施例１と同様に画像や音声の入出力のみが行われる。そして、それら画像や音声の認識合成処理は、実施例１と同様にサーバ１０２Ａ，１０２Ｂで行われる。すなわち、クライアントからネットワークを通じて転送されてくる画像データや音声データに対する処理が、２台のサーバで分散して行われる。

【００３５】サーバ間におけるデータ処理の分散は、各サーバにおける処理状況，すなわち負荷状況や可動状況をクライアント側に報告し、クライアント側で負荷の軽いサーバを選択してデータを転送するなどの方法で可能である。例えば、サーバ１０２Ａ，１０２Ｂに処理状況報告部３００がそれぞれ設けられる。一方、クライアント１００には処理状況判断部３０２が設けられる。データの送り元のクライアント１００では、データ転送前に、まず処理状況報告部３００から各サーバ１０２Ａ，１０２Ｂにおける処理状況の報告を受ける。クライアント１００では、処理状況判断部３０２で各サーバ１０２Ａ，１０２Ｂの処理状況を調べ、負荷の軽いサーバを判断する。そして、負荷の軽いサーバを選択してデータを転送し、その認識や合成を要求する。

【００３６】このように、本実施例によれば、サーバの台数はを増えるものの、処理を複数のサーバに分散させることができ、１つのサーバの負担を軽減することができる。このため、能力的に劣るハードウエア資源を有効に活用することが可能となり、全体としてみればコストダウンを図ることができる。なお、本例において、ハー

ドウエアの一部，例えばデータを蓄積するメモリを各サーバに共通に設けるようにすれば、更に資源の有効活用を図ることができる。

【００３７】

【実施例４】次に、図２（D）を参照しながら実施例４について説明する。この例は、上述した実施例２と実施例３を組み合わせたものである。すなわち、クライアント１００Ｃ，１００Ｄ，１００Ｅと、サーバ１０２Ｃ，１０２Ｄがネットワーク１０４に接続された構成となっている。クライアント１００Ｃ，１００Ｄ，１００Ｅには、識別データ付加部２００，処理状況判断部３０２がそれぞれ設けられている。サーバ１０２Ｃ，１０２Ｄには、クライアント識別部２０２，処理状況報告部３００がそれぞれ設けられている。

【００３８】サーバ１０２Ｃ，１０２Ｄは、クライアント１００Ｃ，１００Ｄ，１００Ｅに対して処理状況を報告する。クライアント１００Ｃ，１００Ｄ，１００Ｅは、サーバ側の処理状況に応じて負荷の軽いものを選択し、自己の識別データを付加して処理対象となるデータを転送する。データの転送を受けたサーバでは、その処理を行い、処理後のデータを該当するクライアントに転送する。

【００３９】このように、本実施例によれば、ネットワーク上に複数のサーバ及び複数のクライアントが用意される。そして、複数のクライアントでは、複数のサーバから負荷の軽いものが任意に選択されて処理が要求される。サーバ側では分散処理が行われるので、大規模なネットワークとなっても柔軟に対応することが可能となって、ハードウエア資源が更に有効に活用されるようになる。

【００４０】

【他の実施例】この発明には数多くの実施の形態があり、以上の開示に基づいて多様に改変することが可能である。例えば、次のようなものも含まれる。

【００４１】（1）システムを構成するカメラや表示装置などの各要素としては、各種のものが知られており、いずれを用いてもよい。例えば、センサ１０としては、赤外線センサ，振動センサ，音声センサなどが使用できる。画像や音声の認識や合成のアルゴリズムも、何ら上記実施例に限定されるものでなく、各種の手法を用いてよい。また、認識や合成以外の処理を行うようにしてもよい。ネットワークに接続するクライアントやサーバの数も、必要に応じて適宜増減してよい。

【００４２】（2）前記実施例は、人物の接近を検知して顔面や音声を認識するようなシステムの場合を例として説明したが、画像及び音声に対して何らかの処理を行うようなシステムであれば、どのようなものにも適用可能である。

【００４３】

【発明の効果】以上説明したように、本発明によれば、次のような効果がある。

（1）システムを画像や音声の入出力を行うクライアントと、音声データや画像データの認識・合成処理を行うサーバとに分離することとしたので、空間的な制約が低減される。

（2）複数のクライアントがネットワークを通じてサーバを共有することとしたので、ハードウエア資源の有効活用を図ることができ、コストの低減が可能となる。

（3）複数のサーバによって分散処理することとしたので、各サーバの負担が低減され、能力的に劣るハードウエアを有効に活用することができる。

【図面の簡単な説明】

【図１】この発明の実施例１の構成を示すブロック図である。

【図２】この発明の実施例の主要構成を示すブロック図である。

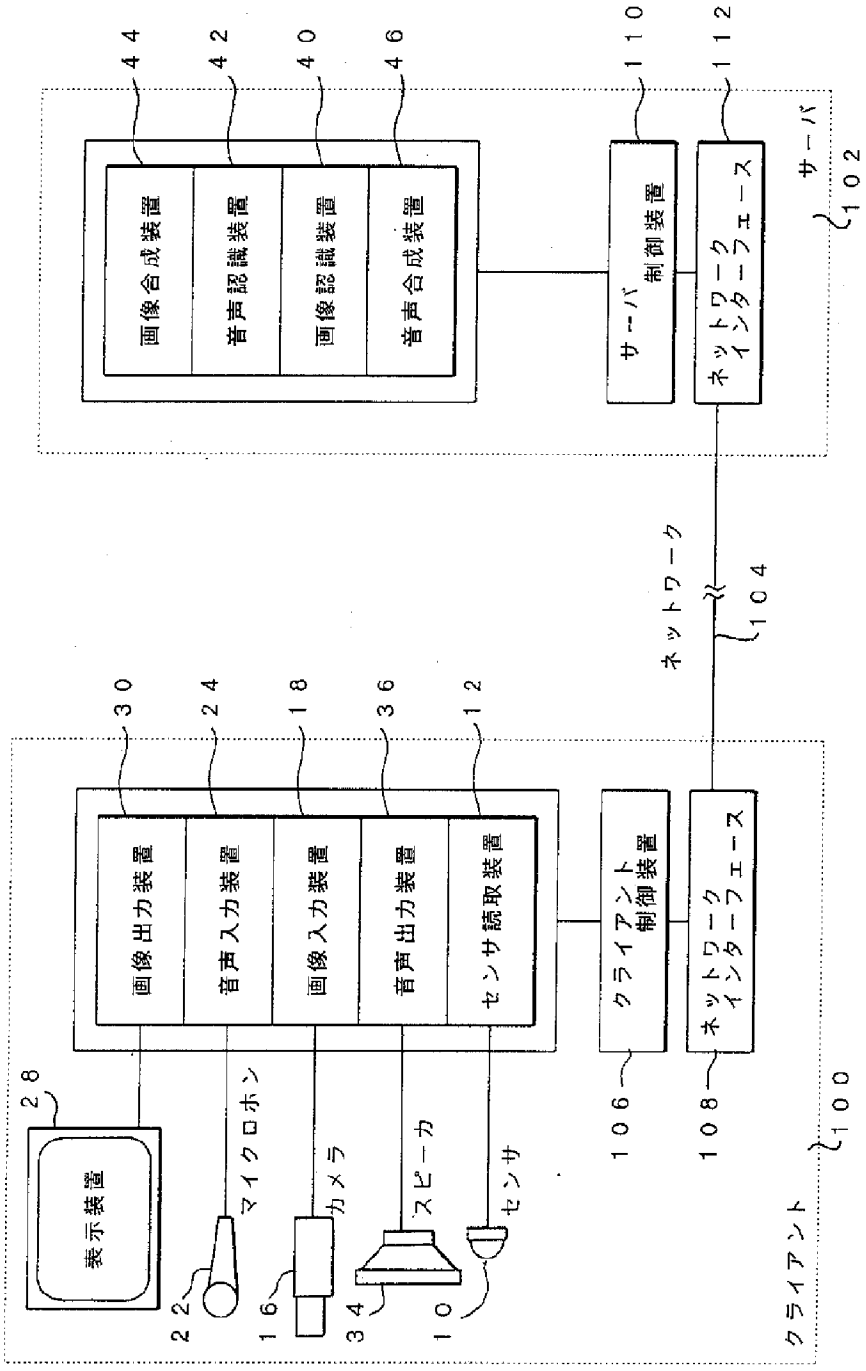
【図３】従来のシステムの一例を示すブロック図である。

【図４】図３の背景技術を複数用いた例を示すブロック図である。

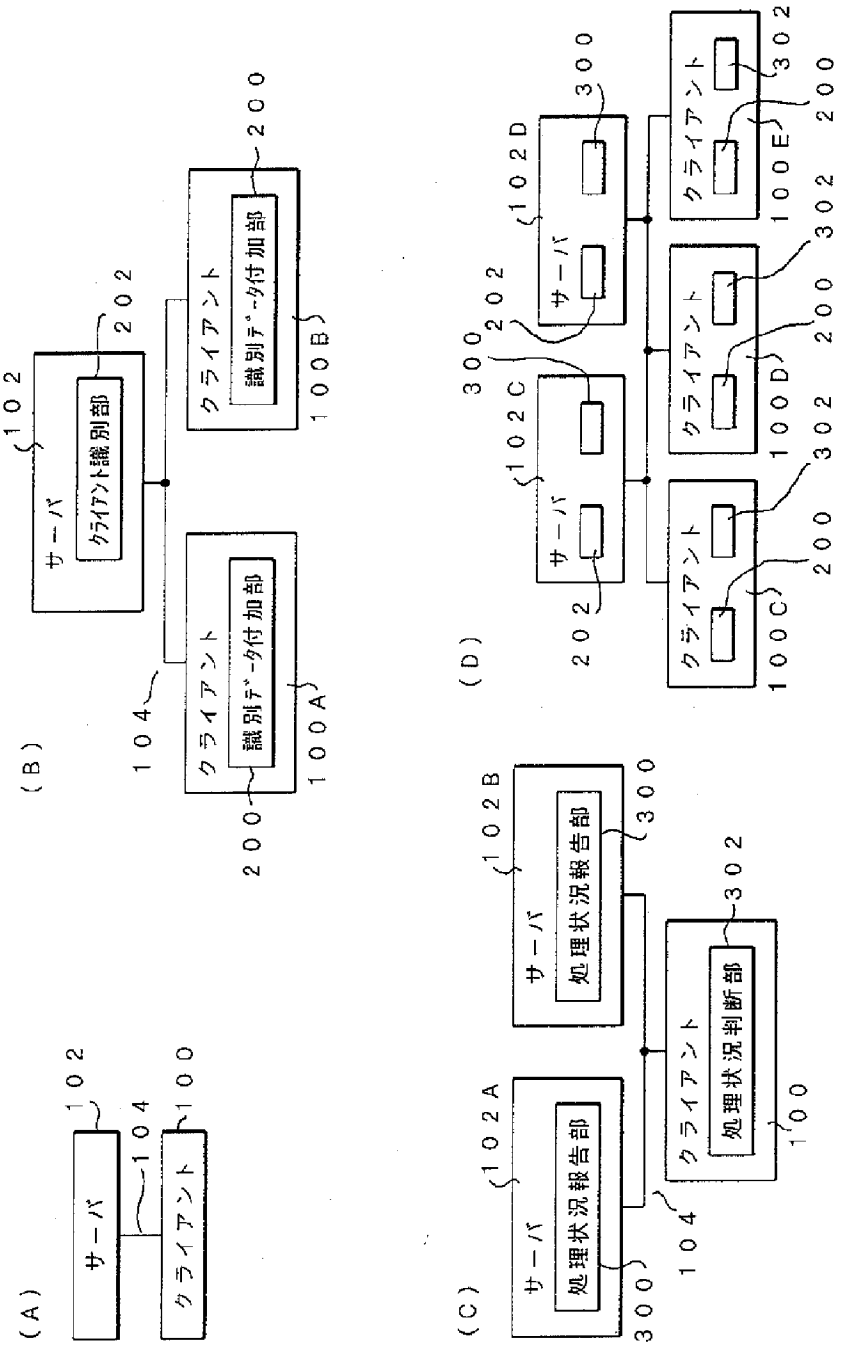
【符号の説明】

１０…センサ
１２…センサ読取装置
１４…センサ読取部
１６…カメラ
１８…画像入力装置
２０…画像入力部
２２…マイクロホン
２４…音声入力装置
２６…音声入力部
２８…表示装置
３０…画像出力装置
３２…画像出力部
３４…スピーカ
３６…音声出力装置
３８…音声出力部
４０…画像認識装置
４２…音声認識装置
４４…画像合成装置
４６…音声合成装置
４８，１０６，１１０…制御装置
１００…クライアント
１０２…サーバ
１０４…ネットワーク
１０８，１１２…ネットワークインターフェース
２００…識別データ付加部
２０２…クライアント識別部
３００…処理状況報告部
３０２…処理状況判断部

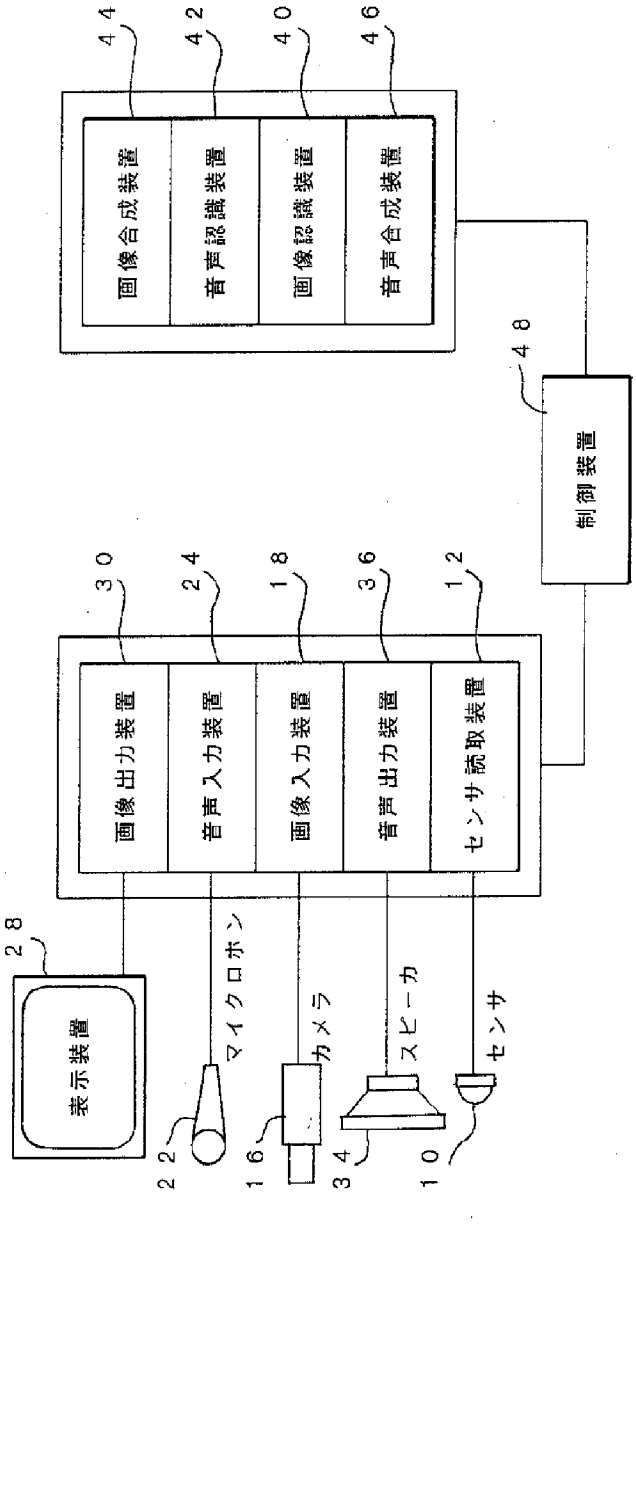
【図１】


サーバ
１０２
画像合成装置
４４
音声認識装置
４２
画像認識装置
４０
音声合成装置
４６
サーバ制御装置
１１０
ネットワークインターフェース
１１２
ネットワーク
１０４
クライアント
１００
画像出力装置
３０
音声入力装置
２４
画像入力装置
１８
音声出力装置
３６
センサ読取装置
１２
クライアント制御装置
１０６
ネットワークインターフェース
１０８
表示装置
２８
マイクロホン
２２
カメラ
１６
スピーカ
３４
センサ
１０

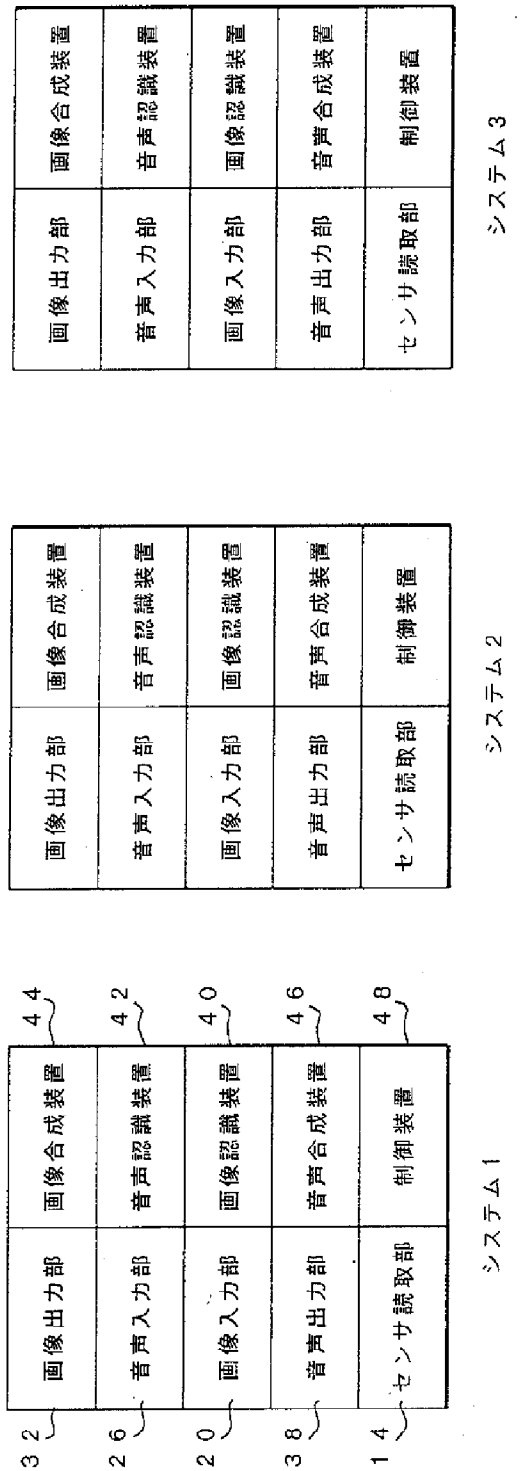
【図２】


(A)
サーバ
102
104
クライアント
100
(B)
サーバ
クライアント識別部
102
202
104
クライアント
識別データ付加部
100A
200
クライアント
識別データ付加部
100B
200
(C)
102A
サーバ
処理状況報告部
300
102B
サーバ
処理状況報告部
300
104
クライアント
処理状況判断部
302
100
(D)
300
202
102C
サーバ
202
クライアント
100C
200
302
102D
サーバ
300
クライアント
100D
200
302
クライアント
100E
200
302

【図３】


４４
４２
４０
４６
画像合成装置
音声認識装置
画像認識装置
音声合成装置
４８
制御装置
３０
２４
１８
３６
１２
画像出力装置
音声入力装置
画像入力装置
音声出力装置
センサ読取装置
２８
表示装置
マイクロホン
カメラ
スピーカ
センサ
２２
１６
３４
１０

【図４】


画像出力部
音声入力部
画像入力部
音声出力部
センサ読取部
画像合成装置
音声認識装置
画像認識装置
音声合成装置
制御装置
システム３
システム２
システム１
３２
２６
２０
３８
１４
４４
４２
４０
４６
４８

フロントページの続き

(51) Int.Cl.[6]	識別記号	ＦＩ	
Ｇ１０Ｌ　3/00		Ｇ１０Ｌ　3/00	Ｒ
			Ｎ
	５５１		５５１Ｚ